

# 基于 Müller–Brown 势的神经网络建模

杨远青 22300190015 CompPhys 24

2025 年 3 月 30 日

## 1 问题描述

### 1.1 函数定义

Müller–Brown 势能函数解析式为：

$$U(x_1, x_2) = s \cdot \sum_{k=1}^4 A_k \exp [\alpha_k(x_1 - a_k)^2 + \beta_k(x_1 - a_k)(x_2 - b_k) + \gamma_k(x_2 - b_k)^2]$$

### 1.2 参数系统

振幅系数:  $\mathbf{A} = (-200, -100, -170, 15)$

二次项参数:  $\boldsymbol{\alpha} = (-1, -1, -6.5, 0.7)$

$\boldsymbol{\beta} = (0, 0, 11, 0.6)$

$\boldsymbol{\gamma} = (-10, -10, -6.5, 0.7)$

中心坐标:  $\mathbf{a} = (1, 0, -0.5, -1)$

$\mathbf{b} = (0, 0.5, 1.5, 1)$

缩放因子:  $s = 0.05$

定义域:  $x_1 \in (-1.5, 1.5), x_2 \in (-0.5, 2)$

势能截断:  $U \leq U_{\text{cut}} = 9$

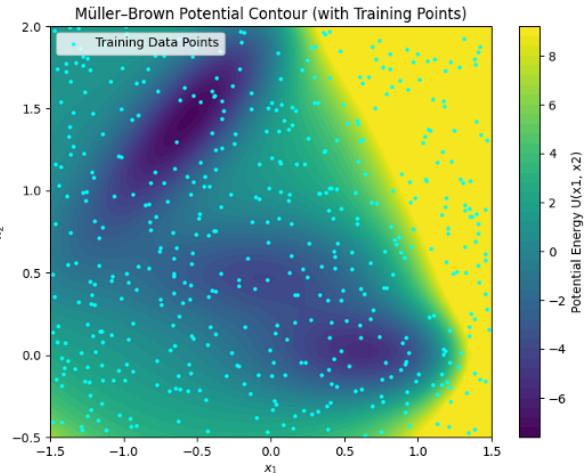


图 1: 待拟合势能面与训练集可视化

### 1.3 评估标准

MAE<sup>1</sup>:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |U(x_1^{(i)}, x_2^{(i)}) - U_{\text{NN}}(x_1^{(i)}, x_2^{(i)})|$$

及训练时间等。

<sup>1</sup>最终 MAE 系列评价指标分为：全局 MAE、训练集 MAE、测试集 MAE，其中全局 MAE 不是训练集和测试集的 MAE 的加权平均值，而是在全势能面上采集的（ $100 \times 100$ ）个点（相当于更大的超测试集），与受限于优化采点策略的训练集、测试集，略有不同。

## 2 模型设计



图 2: 三种神经网络模型架构对比

### 2.1 MLPNet

$$\begin{aligned}
 \mathbf{h}_1 &= \text{ReLU}^\dagger(\mathbf{W}_1 \mathbf{x} + \mathbf{b}_1) \in \mathbb{R}^{128} \\
 \mathbf{h}_2 &= \text{ReLU}(\mathbf{W}_2 \mathbf{h}_1 + \mathbf{b}_2) \in \mathbb{R}^{128} \\
 \mathbf{h}_3 &= \text{ReLU}(\mathbf{W}_3 \mathbf{h}_2 + \mathbf{b}_3) \in \mathbb{R}^{64} \\
 U_{\text{NN}}(\mathbf{x}) &= \mathbf{W}_4 \mathbf{h}_3 + \mathbf{b}_4 \in \mathbb{R}
 \end{aligned}$$

结构特征:

- **基础架构:** 输入 (2) → 隐藏层 (128) → 隐藏层 (128) → 隐藏层 (64) → 输出 (1)
- **激活函数:** ReLU (修正线性单元,  $\text{ReLU}(x) = \max(0, x)$ )
- **信息流动:** 层间全连接, 严格单向传播

## 2.2 StdNet

$$\begin{aligned}
\mathbf{x}_{\text{BN}} &= \text{BatchNorm}^{\ddagger}(\mathbf{x}) \\
\mathbf{h}_1 &= \text{SiLU}^*(\text{BN}_1(\mathbf{W}_1 \mathbf{x}_{\text{BN}} + \mathbf{b}_1)) \in \mathbb{R}^{64} \\
\mathbf{h}_2 &= \text{SiLU}(\text{BN}_2(\mathbf{W}_2 \mathbf{h}_1 + \mathbf{b}_2)) \in \mathbb{R}^{128} \\
\mathbf{h}_3 &= \text{SiLU}(\text{BN}_3(\mathbf{W}_3 \mathbf{h}_2 + \mathbf{b}_3)) \in \mathbb{R}^{64} \\
U_{\text{NN}}(\mathbf{x}) &= \mathbf{W}_4 \mathbf{h}_3 + \mathbf{b}_4 \in \mathbb{R}
\end{aligned}$$

结构特征：

- 规范化设计：引入批归一化 (BatchNorm $^{\ddagger}$ :  $\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}}$ )
- 激活函数：SiLU (Sigmoid 线性单元,  $\text{SiLU}(x) = x \cdot \sigma(x)$ )
- 层级结构：64→128→64 的沙漏型设计，增强特征提取能力

## 2.3 ResNet

$$\begin{aligned}
\mathbf{x}_{\text{BN}} &= \text{BatchNorm}(\mathbf{x}) \\
\mathbf{h}_1 &= \text{ResBlock}_1(\mathbf{x}_{\text{BN}}) \in \mathbb{R}^{64} \\
\mathbf{h}_2 &= \text{ResBlock}_2(\mathbf{h}_1) \in \mathbb{R}^{64} \\
\mathbf{h}_3 &= \text{ResBlock}_3(\mathbf{h}_2) \in \mathbb{R}^{64} \\
U_{\text{NN}}(\mathbf{x}) &= \mathbf{W}_4 \mathbf{h}_3 + \mathbf{b}_4 \in \mathbb{R}
\end{aligned}$$

残差块定义 ( $i = 1, 2, 3$ )：

$$\begin{aligned}
\text{ResBlock}_i(\mathbf{z}) &= \mathbf{z} + \mathcal{F}_i(\mathbf{z}) \\
\mathcal{F}_i(\mathbf{z}) &= \mathbf{W}_{i,2} \cdot \text{SiLU}(\text{BN}_{i,1}(\mathbf{W}_{i,1} \mathbf{z} + \mathbf{b}_{i,1})) + \mathbf{b}_{i,2}
\end{aligned}$$

结构特征：

- 残差连接：通过跳跃连接 ( $\mathbf{z} + \mathcal{F}(\mathbf{z})$ ) 缓解梯度消失
- 恒等映射：保持特征维度一致 (64 维)，避免投影捷径
- 复合结构：每个残差块包含两个线性层，配合 BN 和 SiLU 激活

### 3 训练过程

#### 3.1 训练设置:

- **优化器:** Adam (初始学习率  $1 \times 10^{-3}$ )
- **批量大小:** 64
- **学习率调度:** ReduceLROnPlateau (当验证损失停滞时降低为当前值的一半)
- **早停机制:** 连续 `patience`<sup>2</sup> 个 epoch 验证损失无改善则终止训练

#### 3.2 训练策略:

- **第一阶段:** 原始 500 数据点 (80% 训练集 + 20% 验证集)
- **第二阶段:** 增广至 2500 数据点 (新增 2000 个非均匀采样点)
- **全局超测试集:** 固定 10000 点网格数据 ( $x_1 \in [-1.5, 1.5]$ ,  $x_2 \in [-0.5, 2]$  均匀采样)

#### 3.3 过程对比:

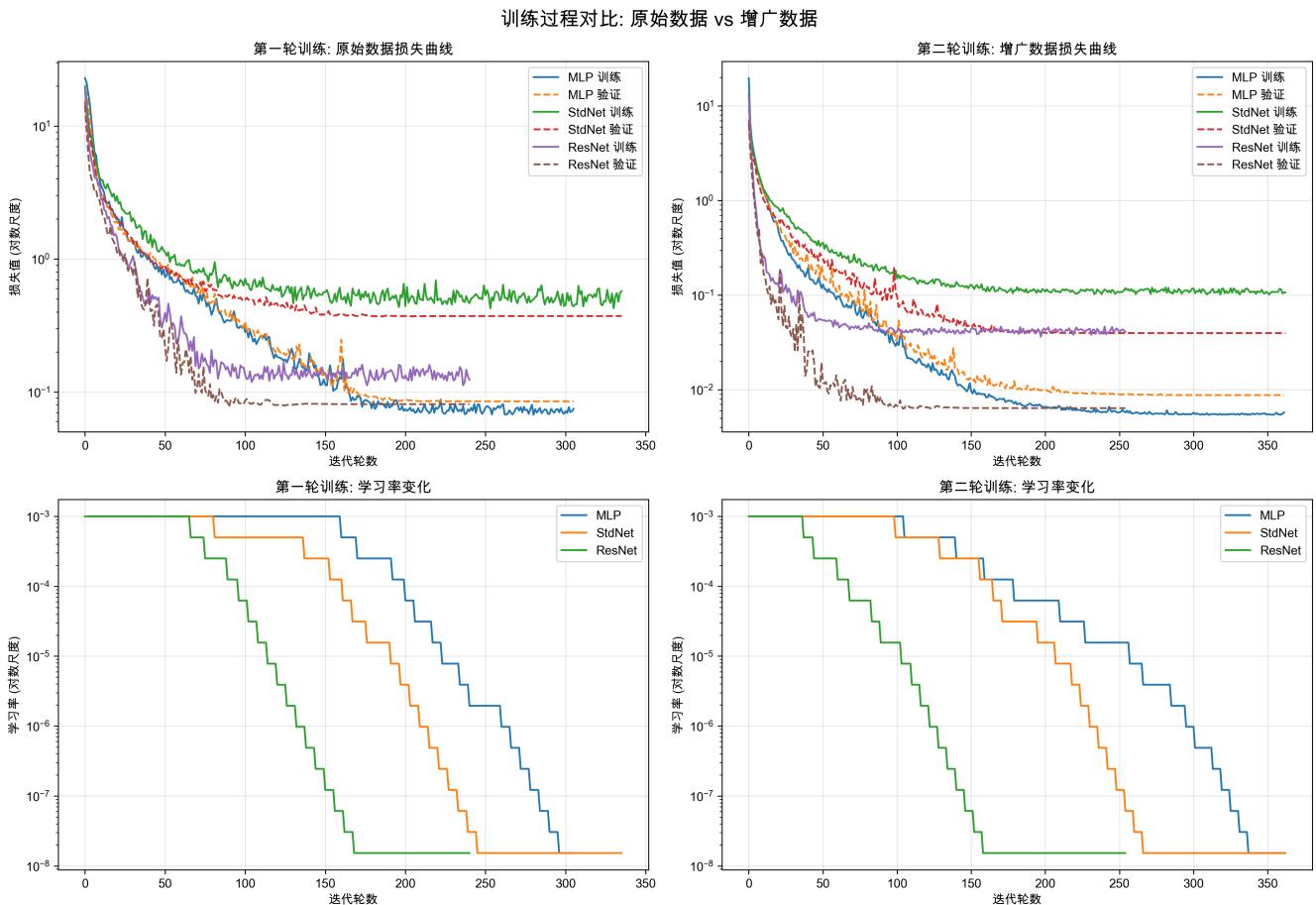


图 3: 两阶段训练过程对比 (左: 原始数据训练, 右: 增广数据训练)

<sup>2</sup>为控制训练时间大致相同, 三个模型的 `patience` 分别设置为 50, 150, 150

## 4 训练结果

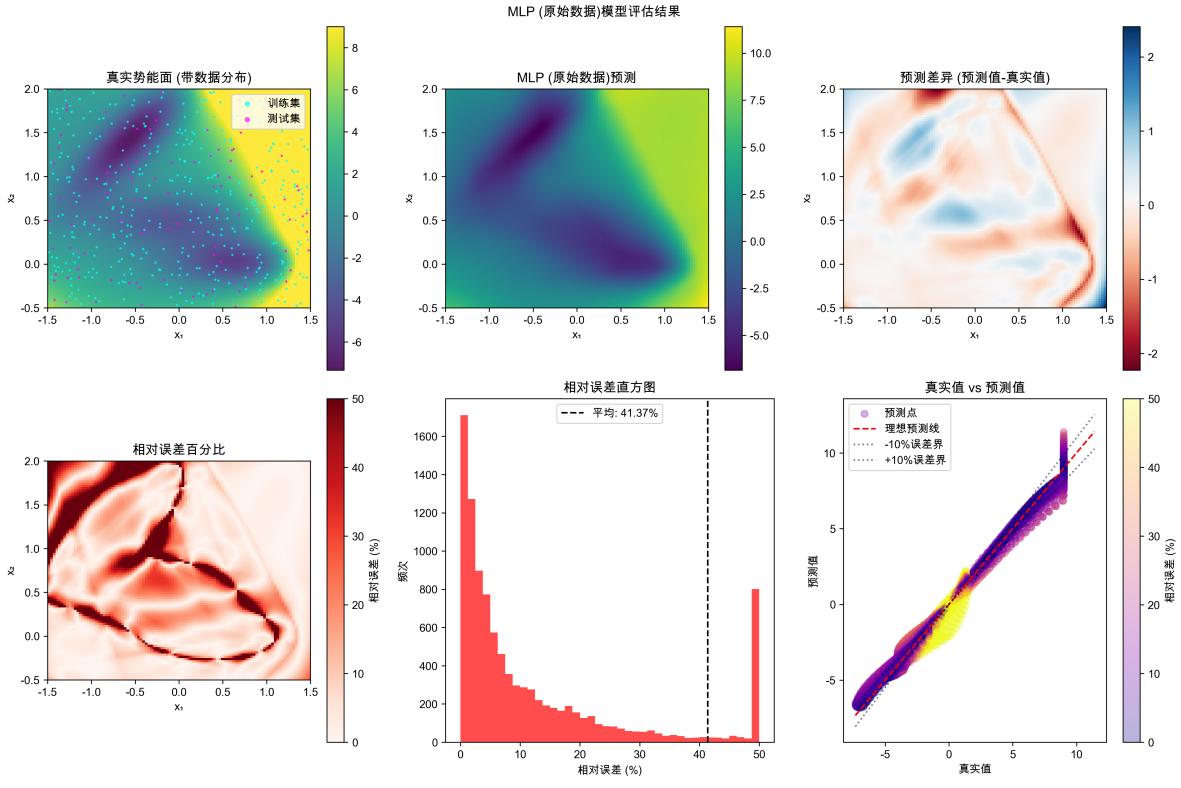
### 4.1 训练结果可视化

参见图4、图5和图6

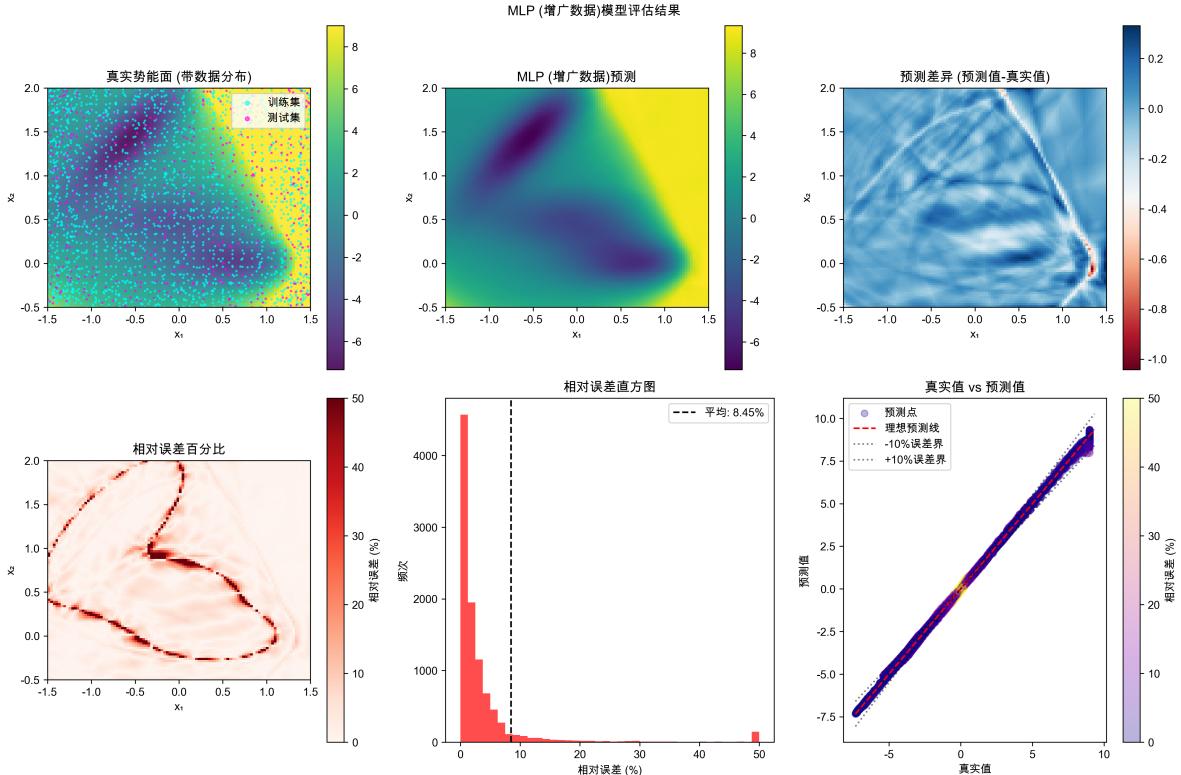
### 4.2 结果分析

从可视化图中可以观察到几个关键现象：

1. **数据增广效果**: 所有模型在使用增广数据后，预测精度显著提高，尤其在势能面的边缘区域。
2. **误差分布特征**: 相对误差主要集中在势能值的过渡区域，而在势能的主要极小值区域（Müller-Brown 势能的三个主要井区）预测精度较高。
3. **相对误差特点**: 所有模型的相对误差分布都呈现右偏分布特性，即大多数区域误差较小（分布直方图左侧集中），少数区域存在较大误差（分布直方图右侧拖尾）。

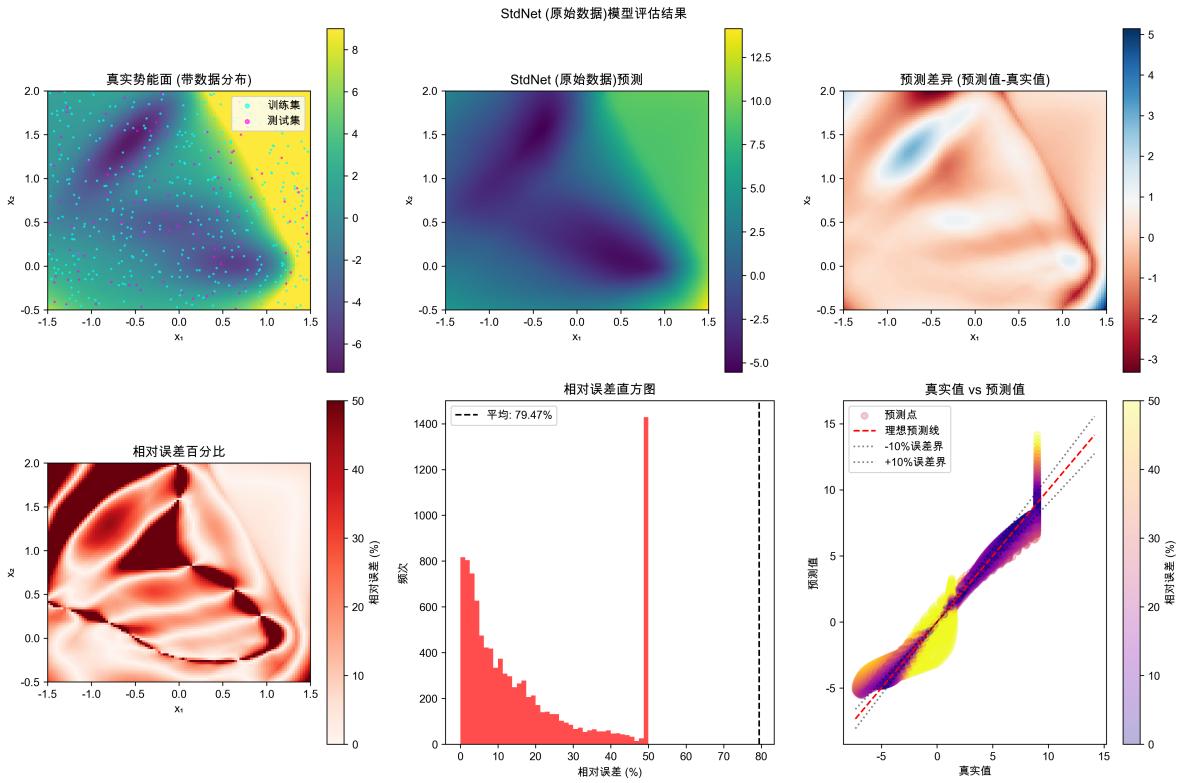


(a) MLP (原始数据)

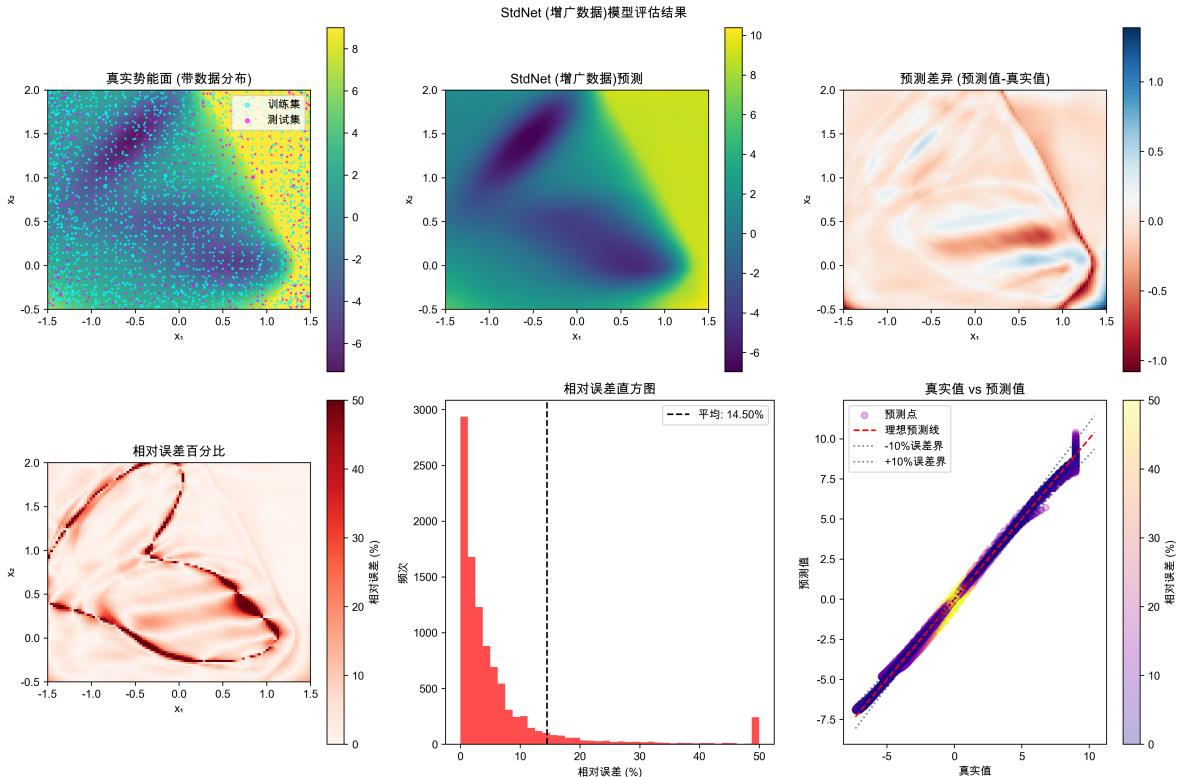


(b) MLP (增广数据)

图 4: MLP 模型在原始数据和增广数据上的训练结果对比

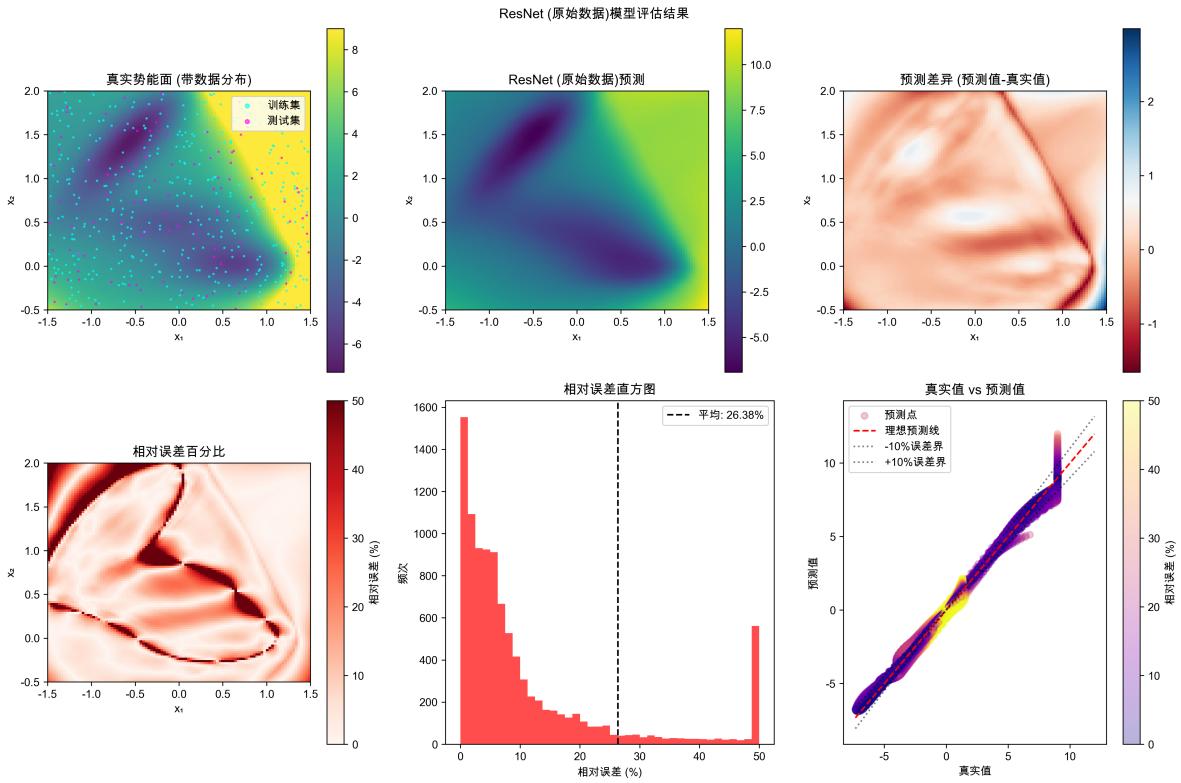


(a) StdNet (原始数据)

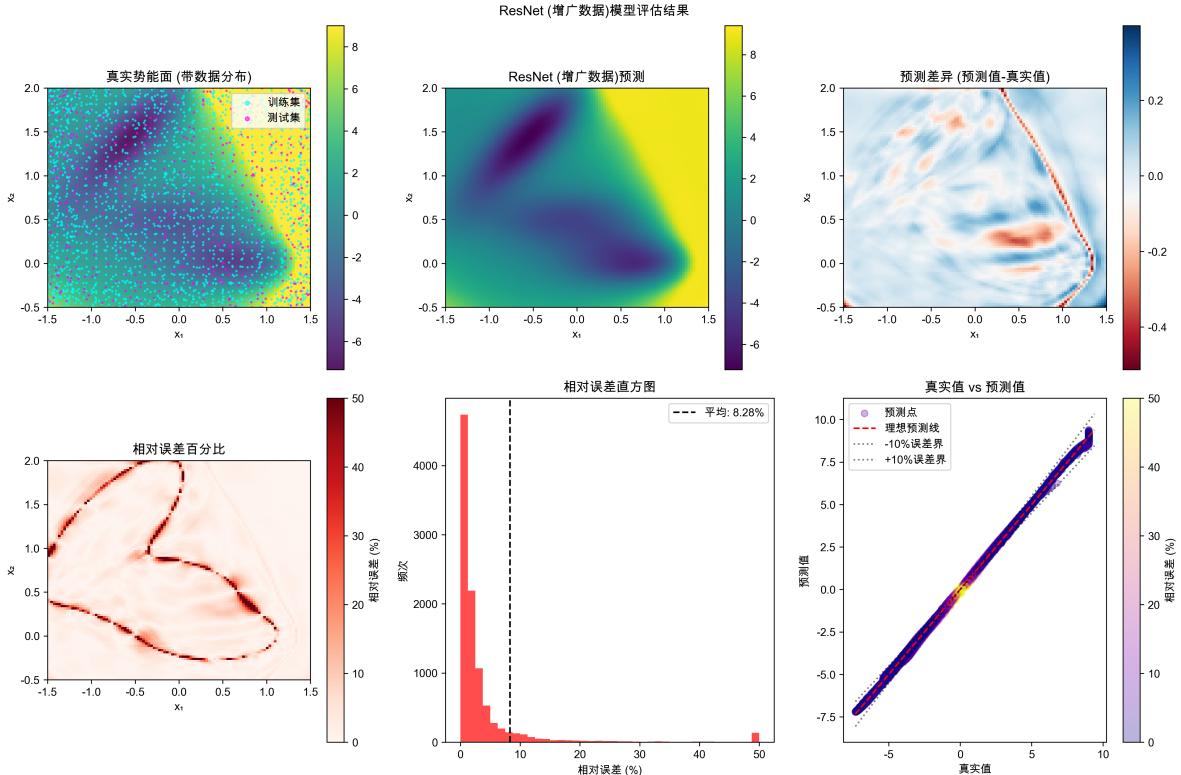


(b) StdNet (增广数据)

图 5: StdNet 模型在原始数据和增广数据上的训练结果对比



(a) ResNet (原始数据)



(b) ResNet (增广数据)

图 6: ResNet 模型在原始数据和增广数据上的训练结果对比

## 5 模型评估

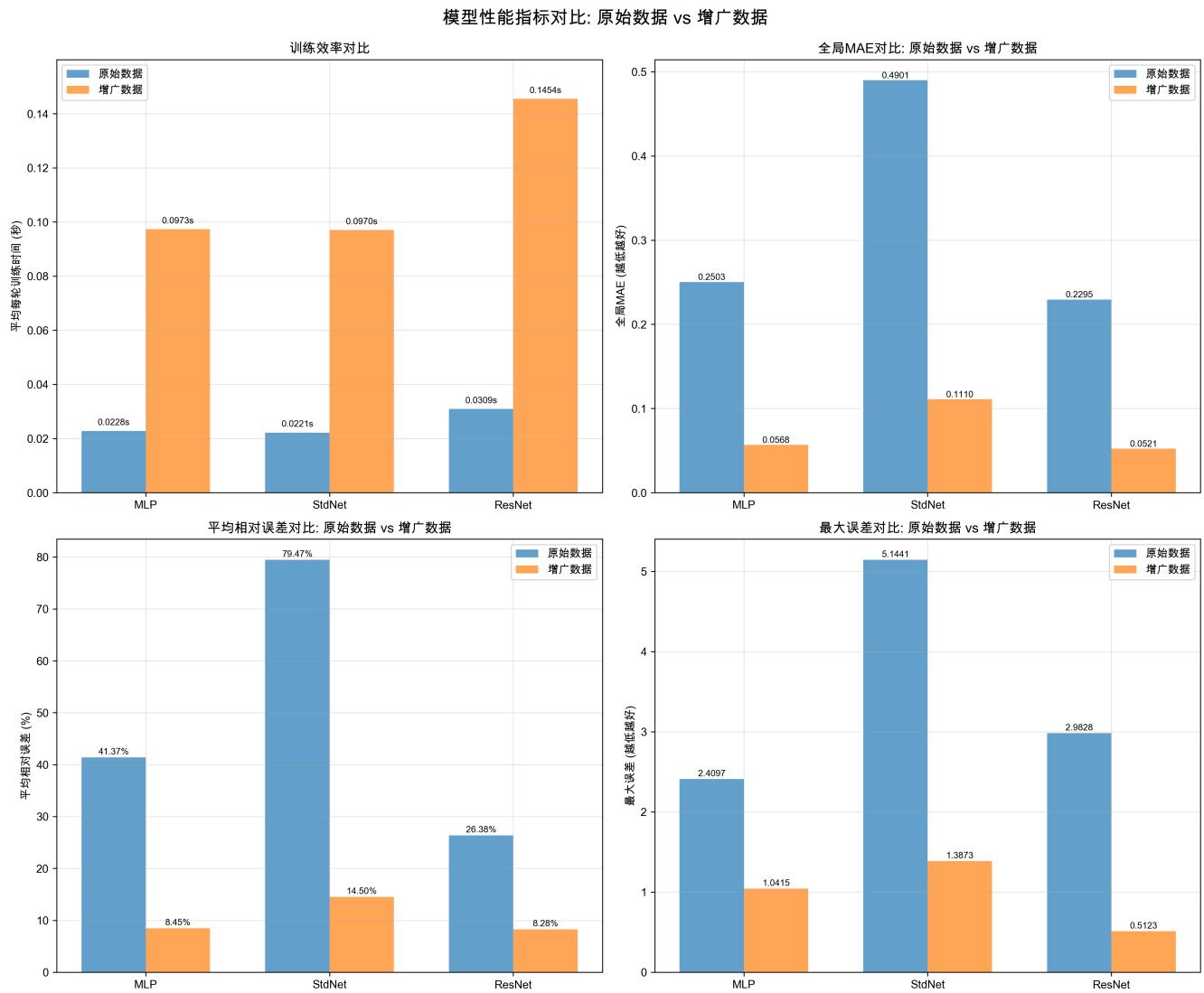


图 7: 各模型在原始数据和增广数据上的性能指标对比

Scaling Law 大展神威, StdNet 表现不尽人意。

## 5.1 详细评估指标

表 1: 模型基本特性对比

指标	MLP	StdNet	ResNet
模型参数量	25,217	33,793	66,561
标准化层	否	是	是
Dropout 正则化	否	是	是
残差连接	否	否	是

表 2: 原始数据上 (500 点) 的训练结果

指标	MLP	StdNet	ResNet
训练轮数	306	336	241
总训练时间 (秒)	6.97	7.44	7.44
平均每轮时间 (秒)	0.0228	0.0221	0.0309
最终学习率	1.53e-08	1.53e-08	1.53e-08
全局 MAE	0.2503	0.4901	0.2295
训练集 MAE	0.1852	0.4008	0.1774
测试集 MAE	0.2095	0.4368	0.2062
RMSE	0.3793	0.7310	0.3431
最大误差	2.4097	5.1441	2.9828
平均相对误差 (%)	41.3736	79.4676	26.3759

表 3: 增广数据上 (2500 点) 的训练结果

指标	MLP	StdNet	ResNet
训练轮数	362	363	255
总训练时间 (秒)	35.22	35.22	37.09
平均每轮时间 (秒)	0.0973	0.0970	0.1454
最终学习率	1.53e-08	1.53e-08	1.53e-08
全局 MAE	0.0568	0.1110	0.0521
训练集 MAE	0.0481	0.1019	0.0452
测试集 MAE	0.0578	0.1176	0.0520
RMSE	0.0874	0.1748	0.0766
最大误差	1.0415	1.3873	0.5123
平均相对误差 (%)	8.4511	14.4967	8.2789

表 4: 数据增广带来的性能改进

模型	全局 MAE 改进	相对误差改进
MLP	77.28%	79.57%
StdNet	77.35%	81.76%
ResNet	77.28%	68.61%